Chapter 5

Error Estimates for the Finite Element Method

5.1 Introduction

In this chapter, we shall focus on two types of error estimates for the finite element method, a priori and a posteriori estimates. (We have previously introduced these terms in the context of quadrature rules.) These two types of estimates serve very different purposes. The main feature of a priori estimates is that they tell us the order of convergence of a given finite element method, that is, they tell us that the finite element error $\|u - u_h\|$ in some norm $\|\cdot\|$ is $O(h^\alpha)$, where h is the (maximum) mesh size and α is a positive integer. The constant in the $O(h^\alpha)$ is generally unknown and is often not of great interest. The goal of these estimates is to give us a reasonable measure of the efficiency of a given method by telling us how fast the error decreases as we decrease the mesh size. In contrast, a posteriori estimates use the computed solution u_h in order to give us an estimate of the form $\|u - u_h\| \leq \epsilon$, where ϵ is simply a number. These estimates accomplish two main goals. First, they are able to give us a much better idea of the actual error in a given finite element computation than are a priori estimates. Secondly, they can be used to perform *adaptive mesh refinement*. In adaptive mesh refinement, a posteriori error estimators are used to indicate where the error is particularly high, and more mesh intervals are then placed in those locations. A new finite element solution is computed, and the process is repeated until a satisfactory error tolerance is reached.

5.2 A priori error estimates

Our goal is to find bounds for the error $u - u_h$ in the finite element approximation of the solution u to our general two-point boundary value problem. We first consider how we should measure the error, that is, what norm we should use. We have already demonstrated that the energy norm $\|\cdot\|_{eng}$ is convenient in one sense because $\|u - u_h\|_{eng} = \min_{\chi \in S_h} \|u - \chi\|_{eng}$. However, it has a few key disadvantages. First, it is highly connected with our problem; in particular, it depends upon the coefficients a, b, and c. This can be a problem if these coefficients vary widely in size. Also, if we were to use this norm as our error measure, we would have to compute $\min_{\chi \in S_h} \|u - \chi\|_{eng}$ for each

choice of a, b, and c. A second disadvantage of the energy norm is that it measures both the error $u - u_h$ in the function values and the error $(u - u_h)'$ in the first derivative. However, sometimes we wish to measure only the error $u - u_h$ in the function values, and in fact this error converges at a higher rate than does the error in the derivative (or more precisely, the error in the H^1 or energy norms). Finally, the energy norm measures an average error in some sense; more precisely, it measures weighted integrals of the squares of the errors over the interval $[0, 1]$. However, controlling the average error leaves open the possibility that the error is very large on a very small portion of the interval $[0, 1]$. Thus we could seek to measure the maximum of the error over the interval $[0, 1]$, i.e., $\|u - u_h\|_{L_\infty([0,1])}$.

We shall present error estimates in three different norms, all of which overcome at least one of the objections to the energy norms presented above. The first option is the H^1 norm, defined as $\|u - u_h\|_{H^1} = (\int_0^1 (u' - u_h')^2\,dx + \int_0^1 (u - u_h)^2\,dx)^{1/2}$. (In all of our norm notation in this section, we shall assume that the norm is being taken over the interval $[0, 1]$ unless otherwise noted.) Note that if $a \equiv c \equiv 1$ and $b \equiv 0$, the H^1 norm is precisely the energy norm. The H^1 and energy norms provide similar measures of the error under many circumstances, but the H^1 norm is independent of the problem under study and is thus more convenient for many purposes.

A second option for measuring the error is to use one of the L_p norms, given by $\|u - u_h\|_{L_p} = (\int_0^1 (u - u_h)^p\,dx)^{1/p}$, $1 \le p < \infty$. Typically it is easiest to choose $p = 2$, but other choices of p sometimes are convenient or necessary. These norms measure only the error in function values and not in derivatives. The third option, the L_∞ norm, also measure only function values, but it measures the maximum $\|u - u_h\|_{L_\infty}$ of the error over the interval $[0, 1]$ instead of measuring an average over this interval as the L_p norms do (again, we may interpret an integral as an average).

We next give error estimates for the piecewise linear and Hermite cubic finite element methods in the various norms listed above. We shall only prove the estimate for the Hermite cubic method in the H^1 norm. This is the simplest of the estimates to prove, and even its proof is somewhat lengthy.

Theorem 5.2.1 *Assume that $0 < a_0 \le a(x) \le A_0$ and $0 < c_0 \le c(x) \le C_0$ for $0 \le x \le 1$, and assume that u satisfies $\mathcal{L}(u, v) = (f, v)$, $v \in \mathcal{A}$ (that is, assume that u satisfies a two-point boundary value problem with homogeneous Neumann boundary conditions). Let also u_h be the Hermite cubic spline finite element approximation to u on a mesh with maximum element size h. Then*

$$\|u - u_h\|_{H^1} \le Ch^3\|u^{(4)}\|_{L_2}, \tag{5.1}$$

$$\|u - u_h\|_{L_p} \le Ch^4\|u^{(4)}\|_{L_p}, \quad 1 \le p < \infty, \tag{5.2}$$

$$\|u - u_h\|_{L_\infty} \le Ch^4\|u^{(4)}\|_{L_\infty}, \tag{5.3}$$

where $u^{(4)}$ is the fourth derivative of u.

Proof. Omitted here for the present.

We next give analagous estimates for the piecewise linear finite element method.

Theorem 5.2.2 *Assume that $0 < a_0 \le a(x) \le A_0$ and $0 < c_0 \le c(x) \le C_0$ for $0 \le x \le 1$, and assume that u satisfies $\mathcal{L}(u, v) = (f, v)$, $v \in \mathcal{A}$ (that is, assume that u satisfies a two-point boundary value problem with homogeneous Neumann boundary conditions). Let also u_h be the piecewise linear finite element approximation to u on a mesh with maximum element size h. Then*

$$\|u - u_h\|_{H^1} \le Ch\|u''\|_{L_2}, \tag{5.4}$$

$$\|u - u_h\|_{L_p} \le Ch^2\|u''\|_{L_p}, \quad 1 \le p < \infty, \tag{5.5}$$

$$\|u - u_h\|_{L_\infty} \le Ch^2\|u''\|_{L_\infty}. \tag{5.6}$$

Proof. Omitted.

We also note that these estimates are valid for a wide range of boundary conditions; we have only presented them for homogeneous Neumann conditions for the sake of simplicity.

As stated previously, our interest in these estimates is largely theoretical: we wish to know how fast our error will decrease as the mesh size decreases. They do provide a very handy basis for comparing methods, however. For example, notice that the error in the Hermite cubic method as measured in the L_∞ norm is $O(h^4)$, while the L_∞ error in the piecewise linear method is $O(h^2)$. Thus the error will decrease much faster when using the Hermite cubic method than when using the piecewise linear method, and the Hermite cubic method is usually more efficient computationally speaking (though perhaps harder to program). There are definitely exceptions, however. The error estimates for the Hermit cubic method require that u have four derivatives, whereas the error estimates for the piecewise linear method only require that u have two derivatives. There are practical situations where u indeed only has two derivatives (perhaps even less), and here it makes much more sense to use the piecewise linear elements. While the Hermite cubics and piecewise linears both yield $O(h^2)$ convergence if u doesn't have more than two derivatives, the Hermite cubics will not be as efficient as the piecewise linears in achieving this accuracy. Thus while the Hermite cubics are very efficient for smooth problems, the piecewise linears will yield optimal convergence for a wider range of problems.

These estimates also give us an excellent tool for dealing with a very practical problem, that of debugging codes. We may observe two things from them. First, there are times when the finite element method should yield an exact solution. For example, assume that the Hermite cubic method is used to approximate a solution u, where $u(x)$ is a cubic polynomial. Since cubic polynomials have a fourth derivative which is 0, the error $u-u_h$ should be 0 (this is true no matter what the coefficients are, so long as they satisfy the conditions given in the theorem). Thus if we set up a problem with a known solution which is a cubic polynomial, the finite element method should return the exact solution. We may also do a rate of convergence test, just as we described in the context of Gaussian quadrature. If such a test yields a suboptimal computed rate of convergence, then there is probably a bug in the code.

5.3 A posteriori error estimation and adaptive mesh refinement

In this section, we shall try to solve the following problem: How should the mesh points $\{x_i\}_{i=0,\ldots,N}$ in a finite element mesh be placed so that a given error tolerance is achieved with maximum efficiency, i.e., with as few mesh points as possible? The answer to this problem depends on many factors, including the coefficients and solution of the given TPBVP, and in fact it is solved in different ways in different circumstances. We give here one solution which is very widely used in practice.

Knowing where to place mesh points before computing a finite element solution is difficult because it generally requires some knowledge about the solution–which is exactly what we don't know since that's what we're trying to compute in the first place. Thus we shall use the following strategy. We will first compute a finite element approximation on a relatively coarse grid (i.e., a grid with relatively few mesh points). We shall use information from this computed solution to guess where more mesh points should be placed, then recompute the finite element solution on the resulting finer mesh. This procedure is continued until a given error tolerance is reached.

To be more precise, let's suppose that we are seeking to find a finite element approximation u_h to the solution u of a TPBVP such that

$$\|u - u_h\|_{L_\infty} \leq tol.$$

Here *tol* might be .01, .001, or something like that. It is difficult to know absolutely for certain that we have reached this goal (unless we know u ahead of time, which we generally don't), so we shall instead settle for being reasonably sure that we are close to it.

We assume here that u_h is the piecewise linear finite element approximation to u. Then (5.6) tells us that $\|u - u_h\|_{L_\infty} \leq Ch^2\|u''\|_{L_\infty}$. Actually, we shall need a sharper version of this estimate, which is

$$\|u - u_h\|_{L_\infty([0,1])} \leq C \max_{1\leq j\leq N} h_j^2 \|u''\|_{L_\infty([x_{j-1},x_j])}.$$

Thus if we can determine C and $\|u''\|_{L_\infty([x_{j-1},x_j])}$ for each j, we will have a reasonable bound on the error as measured in the L_∞ norm.

Theoretical bounds for C are often unrealistic, although sometimes theory can determine C more closely than we have done above. We will discuss two strategies for dealing with C. The first is to simply ignore it. If C isn't too big (say, 2 or 3) and we are somehow able to ensure that $\max_{1\leq j\leq N} h_j^2\|u''\|_{L_\infty([x_{j-1},x_j])} \leq tol$, then in fact $\|u - u_h\|_{L_\infty([0,1])} \leq Ctol$. Although we would prefer not to be off by this factor of C, we generally will be willing to accept it. If it is important to have firm bounds on $\|u - u_h\|_{L_\infty([0,1])}$ (that is, if it is important to know this error precisely and not just up to a constant factor), then well-chosen experiments can yield a reasonable estimate for C. Finite element approximations to a variety of problems with known solutions may be computed on a variety of meshes, both uniform and non-uniform. In each experiment, C is then estimated by $\frac{\|u-u_h\|_{L_\infty([0,1])}}{\max_{1\leq j\leq N} h_j^2\|u''\|_{L_\infty([x_{j-1},x_j])}}$. This will yield a range of possible values for C, and the greatest is then used, perhaps with a little extra added on just to be sure.

We shall expend most of our effort in the estimation of $\|u''\|_{L_\infty([I_j])}$ (here $I_j = [x_{j-1}, x_j]$, $1 \leq j \leq N$). Our goal is to construct an error estimator $EE(I_j)$ so that $h_j^2\|u''\|_{L_\infty([I_j])} \approx EE(I_j)$. Here we shall present two possible choices of the error estimator EE. The first choice of EE is very specific to the problem we are studying. Recall that

$$-(au')' + bu' + cu = f,$$

or

$$-au'' - a'u' + bu' + cu = f.$$

Since we have assumed that $a(x) \geq a_0 > 0$ for all x, we can solve for $u''(x)$ precisely:

$$u''(x) = \frac{f(x) + a'(x)u'(x) - b(x)u'(x) - c(x)u(x)}{a(x)}.$$

We of course don't know $u'(x)$ and $u(x)$, but we have estimates $u_h'(x)$ and $u_h(x)$ for these values. Thus

$$u''(x) \approx \frac{f(x) + a'(x)u_h'(x) - b(x)u_h'(x) - c(x)u_h(x)}{a(x)},$$

and our first error estimator is

$$EE(I_j) = h_i^2 \left\| \frac{f + a'u_h' - bu_h' - cu_h}{a} \right\|_{L_\infty(I_j)}$$

The accuracy of this estimator can be proven by using bounds for $u - u_h$ and $(u - u_h)'$ in L_∞, although we don't do so here. While it is an excellent choice for use with piecewise linear finite element methods for second-order TPBVP's, it relies heavily upon the fact that the order of the differential equation and the order of convergence of the finite element method are the same. This is a fairly severe restriction; for example, this method (or a similar one) would not work when using the Hermite cubics to solve our TPBVP.

A second option is to estimate u'' by second difference quotients of u_h. This approach is more generally applicable than that given above and could be extended to higher-order methods. It is not always rigorously justified theoretically speaking, but generally works well in practice. (We shall consider our use of this estimator to be purely experimental mathematics–try it and see if it works with little heed paid to theory!) We first note that for any mesh point x_j,

$$u''(x_j) = \frac{1}{h_j(\frac{h_j+h_{j+1}}{2})}u(x_{j-1}) - \frac{2}{h_j h_{j+1}}u(x_j) + \frac{1}{h_{j+1}(\frac{h_j+h_{j+1}}{2})}u(x_{j+1}) + O(\max\{h_j, h_{j+1}\}).$$

We shall simply disregard the $O(\max\{h_j, h_{j+1}\})$ and replace $u(x_j)$ with $u_h(x_j)$. Doing so, we make the definition

$$\delta^2 u_h(x_j) = \frac{1}{h_j(\frac{h_j+h_{j+1}}{2})}u_h(x_{j-1}) - \frac{2}{h_j h_{j+1}}u_h(x_j) + \frac{1}{h_{j+1}(\frac{h_j+h_{j+1}}{2})}u_h(x_{j+1}).$$

Note that $\delta^2 u_h$ is defined only at mesh points x_j. In order to estimate u'' over I_j, we shall take an average of $\delta^2 u_h(x_{j-1})$ and $\delta^2 u_h(x_j)$, except at the boundaries:

$$EE(I_j) = \begin{cases} h_1^2|\delta^2 u_h(x_1)| \text{ if } j = 1, \\ h_j^2\frac{|\delta^2 u_h(x_{j-1})+\delta^2 u_h(x_j)|}{2} \text{ if } 2 \le j \le N-1, \quad j = 2, ..., N-1, \\ h_N^2|\delta^2 u_h(x_{N-1})| \text{ if } j = N. \end{cases}$$

We shall now present an algorithm for solving the problem we originally stated in this section, which is: For u solving a TPBVP, find u_h such that $\|u - u_h\|_{L_\infty([0,1])} \le tol$, where tol is given. We shall simply assume that C is 1; if a value for C has been determined, one may replace $EE(I_j)$ in what follows with $CEE(I_j)$. Our algorithm is as follows.

Step 1. Choose a mesh size h and calculate a first approximation u_h^{old} on a uniform mesh of size h. One could choose $h \approx \sqrt{tol}$, for example, though h could be larger.

Step 2. Using u_h^{old}, calculate $EE(I_j)$ for $j = 1, ..., N$. If $EE(I_j) \ge tol$, then subdivide I_j into two new mesh intervals by adding a new mesh point at the center of I_j. If $EE(I_j) < tol$, leave I_j alone.

Step 3. If $EE(I_j) \ge tol$ for at least one j, calculate a new approximation u_h^{new} on the new mesh and replace u_h^{old} with u_h^{new}.

Step 4. Repeat steps 2 and 3 until $EE(I_j) < tol$ for $j = 1, ..., N$.